The Genetic Code—Yesterday, Today, and Tomorrow

F. H. C. CRICK

Medical Research Council Laboratory of Molecular Biology
Cambridge, England

This is an historic occasion. There have been many meetings about the genetic code during the past ten or twelve years but this is the first important one to be held since the code became known. When I came to the States early in 1965 I brought with me tentative allocations for many of the 64 triplets. based mainly on the early work of Leder and Nirenberg, the results from the random polymers and the mutagenesis data. I telephoned Marshall Nirenberg, who told me of his latest allocations. A little later I saw Gobind Khorana and heard the first results he was getting using polymers with repeating sequences. I also visited George Streisinger and was told about the preliminary amino acid sequences due to a phase shift in the phage lysozyme. From all this we were able to work out the meaning of several of the remaining doubtful triplets. By March 1965 the great majority of triplets had been unambiguously identified and just a few remained unallocated. It was a most exciting occasion for me, travelling about the country and seeing how the various lines of evidence fitted together. Leslie Orgel and I were also able to make a shrewd guess at several anticodons, and it was at about this time, too, that the idea of "wobble" was invented.

I thought I would use the occasion of this introduction for some reminiscences about the past, some reflections on the present state of the code, and a few words about the future development of the subject.

As we all know, the genetic code is a triplet code of the nonoverlapping type. It is highly degenerate, in a semi-systematic way, most of the 64 triplets standing for one amino acid or another. It is universal, or nearly so. Our present knowledge of it can be neatly summarized by the table given at the beginning of this volume (page 1).

How did ideas about coding start? Of course, the early work of Brachet (1944) and Caspersson (1947) suggested that there was a very intimate connection between RNA and protein synthesis. We know now that this was in some sense misleading, since the RNA which was observed was not in fact messenger RNA but ribosomal RNA. Nevertheless the evidence served to make the idea familiar. In addition there was the famous "one gene—one enzyme" hypothesis of Beadle (1945), but this was never

extended, as far as I know. to embrace the idea of a detailed linear code relating the two sequences. It is usual to quote as the first reference a paper by Caldwell and Hinshelwood. published in 1950. They state their basic idea as follows—"In the synthesis of protein, the nucleic acid, by a process analogous to crystallization. guides the order by which the various amino acids are laid down; in the formation of nucleic acid the converse holds, the protein molecule governing the order in which different nucleotide units are arranged." We now know that, in outline, the first of these ideas is correct, but that the second one is quite erroneous, and in fact contradicts what later came to be called the Central Dogma. They were quite clear, however, that the important thing to do was to specify the sequence of the amino acids. and that the folding of the protein would follow from the sequence. Their ideas about coding were confused, as can be seen from the following quotation: "In a protein, about 23 different amino acids occur, whereas in a nucleic acid only 5 basic units are found—two pyrimidine nucleotides. two purine nucleotides, and ribose phosphate. Clearly, there cannot be a one-to-one correspondence between the position of an individual amino acid in the protein part of a nucleoprotein and the position of an individual nucleotide in a nucleic acid part. If, however, it is assumed that, in the synthesis of a protein at the surface of a nucleic acid polymer, the amino acid side-chain which is guided into a particular place depends on the nature and relative position of two adjacent nucleotide units, the difficulty can be overcome. Twenty-five different internucleotide arrangements are possible, and this is of the right order to give correspondence with the number of different possibilities in the protein chain." As can be seen from this paragraph. Caldwell and Hinshelwood have mistakenly put the ribose phosphate on the same level as the purines and pyrimidines. They did not state whether their code, which was in effect a doublet code. was overlapping or nonoverlapping, but I get the impression that they intended a nonoverlapping one. Although the paper was occasionally referred to, it does not appear to have had any serious influence on later thinking about coding.

A paper which did have some influence was published by Alexander Dounce in 1952. This paper

is really concerned with possible chemical mechanisms for protein synthesis. Dounce also stated quite clearly what we should now call the "sequence hypothesis," and in addition suggested that the (messenger) nucleic acid molecule is concerned in transferring the energy necessary for peptide bond synthesis. He suggested a mechanism by which amino acids were joined to each phosphate of the ribonucleic acid backbone, probably by phosphoamide bonds. Since there was one amino acid for every nucleotide (or, as we should say, the coding ratio was one) he was necessarily led to suggest an overlapping code. The only special feature of this was that the triplet XYZ was given the same amino acid as the triplet ZYX. He correctly pointed out that this gave forty possible groups. However, he did not seem to realize that a code of this sort might be disproved from known sequences, although it should be remembered that at that date very few polypeptide sequences had been worked out. It is interesting to note that the process of recognition of the triplets was by a set of special enzymes, acting at the actual site of the template.

Dounce also suggested a mechanism for synthesizing RNA from RNA, again using four special enzymes, with a diphosphonucleic acid as an intermediate. Dounce's emphasis on the possible chemical mechanisms led to the coding aspect of his paper being somewhat overlooked.

The idea of coding was greatly helped by knowledge of the structure of DNA, published in 1953. Its simplicity excited many people, including the cosmologist, George Gamow. An abbreviated account of Gamow's work first appeared in a short letter to *Nature* in 1954, and this was followed by a longer account in the *Proceedings of the Royal Danish Academy*. I am the proud possessor of one of the early drafts of this paper, then entitled *Protein Synthesis by DNA Molecules*, the authors of which are G. Gamow and C. G. H. Tomkins! (Gamow once told me that he submitted this paper to the *Proceedings of the National Academy* but they rather objected to the imaginary Mr. Tomkins as an author and for this reason it was eventually published by the Royal Danish Academy, although with Gamow as sole author.) The paper is based on the idea that protein synthesis takes place on the surface of double-helical DNA and that the base sequence on the inside of the structure forms a series of cavities, each of which is specific for one of the amino acids. It is not stated in detail how the amino acids recognize these cavities, but the suggestion is fairly clear that they do so by the side chains fitting in stereo-chemically, without any assistance from special enzymes.

Gamow, like Dounce, was concerned that the units in an extended polypeptide chain are separated by only about 3.6 or 3.7 A, and for this reason his code was also of the overlapping type. Concerning the number of amino acids, he says that this "is usually taken as twenty, although actually there may be a few more." In fact, his Table 1 lists 25. The first 20 includes both cystine and cysteic acid (*sic*) and also hydroxyproline, but not asparagine and glutamine. The number is made up to 25 by the inclusion of norvaline, hydroxyglutamic acid, and canine (whatever that is). It was when we first saw this list, I think in the summer of 1953, that Watson and I, sitting in the Eagle at Cambridge, drew up the standard list of twenty which we have today.

The importance of Gamow's work was that it was really an abstract theory of coding, and was not cluttered up with a lot of unnecessary chemical details, although his basic idea that the double-stranded DNA was the template for protein synthesis was, of course, quite wrong. What he did realize clearly was that an overlapping code put restrictions on the amino acid sequences, and that it should be possible to prove, or at least disprove, various overlapping codes by studying known amino acid sequences.

It was at about this time that Gamow founded that strange organization, the RNA Tie Club. This was a club, limited to 20 members (one for each amino acid), of people who were interested in coding problems. It was not a truly representative group of all those in the field, but rather a haphazard collection of Gamow's friends. There were also supposed to be four honorary members, one for each of the four bases, though I do not think that more than two of these were ever elected. The club had a special tie, designed by Gamow and made by a haberdasher in Los Angeles, and each member was supposed to have a tie-pin with the abbreviation for his own amino acid marked on it. I have the tie, but I do not remember ever having had a tie-pin.

A small number of papers was circulated to members of the club and, in particular, in about 1955 I wrote one myself with the title "On Degenerate Templates and the Adaptor Hypothesis." This paper was never published, although parts of it were quoted later in a review by Mahlon Hoagland (1960). It starts off by discussing which amino acids should be included in the standard list. It then goes on to point out that Gamow's detailed scheme could not work for insulin, and that if the insulin data were combined with that for β-corticotropin, a very neat disproof of his code was possible. However, the most fundamental objection to Gamow's scheme was that it did not specify the *direction* of an amino acid sequence, since the DNA

structure itself is not polar. This, of course, was a frequent puzzle in the early days, although it was realized that there might be a special mechanism to determine the direction of reading of the DNA.

What the paper really did was to bring out the general nature of Gamow's ideas. It pointed out that in Gamow's scheme several different triplets could code one amino acid, and it introduced the word degeneracy to describe this. It also emphasized that Gamow's code was an overlapping code, and made the interesting confession: "Watson and I, thinking mainly about codes by hypothetical RNA structures rather than by DNA, did not seriously consider this type of coding." Finally it suggested that Gamow's scheme was essentially abstract, and only paid lip service to structural considerations.

The importance of this unpublished note, however, was that it was the first time that the adaptor hypothesis was put down on paper. I later published a very short remark about it in order to give the idea wider circulation (Crick, 1957). The adaptor was first thought of as a small molecule which was capable of specifically hydrogen bonding with a nucleic acid template. It was not originally conceived to be as big as the present transfer RNA. The role of activating enzymes, however, in providing the specificity for the combination between an amino acid and its adaptor, was clearly stated.

The immediate consequence of the adaptor hypothesis for the coding problem was that all sorts of codes were possible, and it thus became increasingly difficult to disprove any of them. I pointed out that for an overlapping triplet code there cannot be more than 256 different amino acid pairs (out of a possible 400), since any sequence of four base pairs implies a definite pair of amino acids. However, it was left to Brenner (1957) to make this test more efficient, and to show that if the code were universal, the sequences then known made all fully overlapping triplet codes impossible. Even before this, the outlook for coding was not very promising. My paper ends: "Altogether the position is rather discouraging. Whereas on the one hand the adaptor hypothesis allows one to construct, in theory, codes of bewildering variety, which are very difficult to reject in bulk, the actual sequence data, on the other hand, gives us hardly any hint of regularity, or connectedness, and suggests that all, or almost all sequences may be allowed. In the comparative isolation of Cambridge I must confess that there are times when I have no stomach for decoding."

Of course our basic difficulty, as we now know, was that the code is not overlapping and, consequently, there are few if any restrictions on amino acid sequences. This gave the theoretician of those days no real problem on which to work. I am amused to notice, looking back, that I prefaced my paper with the quotation: "Is there anyone so utterly lost as he that seeks a way where there is no way?"

The next significant contribution to the coding problem turned out to be rather an unfortunate one. This was the idea put forward by Griffith, Orgel, and myself, of a comma-less code (Crick et al. 1957). In such a code some of the triplets stand for amino acids ("make sense") while others do not, and are "nonsense." Read in one phase, the message makes sense everywhere. Read out of phase, in either of the two possible ways, it makes nonsense everywhere, thus automatically solving the problem of how to select the correct phase. The rule to be obeyed is that the overlap triplets formed by reading part of one sense triplet and part of an adjacent sense triplet are always nonsense. To our surprise we found that for a four-letter code the maximum number of sense triplets is 20, and that several codes for 20 could be written down. It was even possible to imagine how to include one or two punctuation marks in addition.

This turned out to be one of those nice ideas which is, nevertheless, completely wrong. We ourselves began to lose faith in it when we eventually noticed the wide range of DNA compositions which occur in various micro-organisms, as shown by Belozersky and Spirin (1958) and by Lee, Wahl, and Barbu (1956). Unfortunately, people found the idea so pretty that it was widely referred to, and even found its way into a popular book on the subject. Personally, I was always very undecided about it, as can be seen by my review for the Society of Experimental Biology in 1958 (Crick, 1958).

By 1959 the coding problem was at a very low ebb. Several people had explained how they hoped to show that the gene and the protein it produced were collinear, but nobody seemed anywhere near doing it. It is one of the curiosities of our subject that it was finally proved, using the expected method, only a little time before the code was finally worked out (Yanofsky et al., 1964). Shortly after this, my colleagues at Cambridge confirmed Yanofsky's result in an unexpected way (Sarabhai, et al., 1964).

In 1959 the sequence hypothesis still seemed highly likely on general grounds, but the detailed evidence on the base composition of DNA and ribosomal RNA made it difficult to produce a convincing scheme. At about this time Sinsheimer (1959) suggested a two-letter rather than a four-letter code, but this was rather a desperate measure. Looking back, we can see clearly that what was missing was the idea of messenger RNA. When the

experimental evidence for this began to come in, the coding problem began to get somewhere. The breakthrough came, as we all know, by the discovery by Nirenberg and Matthaei (1961) that poly U could act as a messenger, and that it coded for polyphenylalanine. Shortly after this, we were able to produce genetic evidence strongly suggesting that the code was a triplet code (Crick et al., 1961). By that time, the evidence from the changes in amino acid sequence found in mutants of human hemoglobin and of Tobacco Mosaic Virus, when added to Brenner's argument, made it highly likely that the code was nonoverlapping.

There is no need to give a detailed account of the history of the last few years, which is, in any case well covered by this volume and by the previous Cold Spring Harbor volume for 1963 (Vol. 28). After the work using messengers of known composition but random sequence, mainly by Nirenberg's and Ochoa's groups, there was a pause, since it was not clear how to go on and find the order of bases within the triplets. This led to a flurry of theoretical papers, most of which are best forgotten. It was of course obvious that the code, although a triplet one, had doublet features, but it was not at all easy to arrive at an accurate version of the code from the data then available. I myself had several shots at it, but did not think they were worth publishing. One of the best guesses is a little-known paper by Rychlik and Šorm (1962), suggesting a predominantly doublet code. To obtain this, they made extensive use of the known replacements of amino acids in proteins. Their code is interesting because much of it has turned out to be approximately correct as far as the doublet aspect is concerned. Otherwise the best guess was by Eck (1963), who suggested that in one place in the triplet U equalled C and A equalled G. However, his actual allocation of triplets was erroneous in several places, even allowing for the fact that he had the bases in the wrong order. We can see now, from the known code, that it would have been almost impossible to have deduced it correctly at that time. Even if the detailed idea of wobble had occurred to anyone, there are too many possible wobble theories to choose from, unless one is helped by the experimental data. Moreover, the evidence itself was not adequate. What finally solved the problem, as we all know, was the triplet binding method of Leder and Nirenberg (1964).

Nevertheless, theory was of some use from time to time as a guide to the experimentalists. It certainly suggested that the code was highly degenerate, at a time when most of those working on the cell-free system thought that perhaps only 30 triplets stood for amino acids. Perhaps more important, all the discussions about coding focused

attention on the problem, and made it more real to people who might otherwise have ignored it.

So much for the past. Now let us turn and look at the situation as it is today. The general position, as can be clearly seen from the articles in this volume, is that the genetic code is known, at least for E. coli, but that it is not known with complete certainty. In particular, some triplets (such as UGA, AUA, and AGC) have been guessed rather than firmly established. One of our main problems, therefore, is to learn how to prove the meaning of a triplet beyond reasonable doubt. Once efficient methods for doing this have been worked out for E. coli, it will easily be possible to check the code for other organisms. Already the excellent agreement between the code deduced for E. coli and the changes in amino acid sequence observed in mutant human hemoglobin and in mutant Tobacco Mosaic Virus strongly suggests that the code is very similar in most organisms, and this is supported by the limited evidence from cell-free systems.

We must therefore examine the possible methods available to us. The method of triplet binding enables one to allocate amino acids to many of the 64 triplets. The difficulty with the method is that it can only be trusted when it gives a clear, strong, and unambiguous result. Unfortunately, there are some triplets that almost certainly code a particular amino acid but do not give a positive result in the binding test. Equally, some triplets show a strong binding to one amino acid but, in addition, a weak binding to one or two others. It is highly likely that most of these weak bindings are simply artifacts of the method, though it is just possible that some of them may be meaningful.

The great advantage of the binding method is that it enables us to study one triplet at a time. The next most useful method appears to be the construction of polymers of known sequences, and in particular of repeating sequence, as Khorana is doing. In this work synthesis is used, not merely binding, and so far it has not produced any result which we have reason to suspect is unreliable. It should therefore be possible to confirm many, if not all, the 64 triplets in this way. In addition, it has given us what is probably the best evidence for the direction of reading. In spite of earlier results to the contrary, it now appears that the amino end of the polypeptide chain corresponds to the 5' end of the mRNA. This has also been confirmed by the phase-shift results and by genetic means.

Both the binding methods and the method of synthesis are open to the objection that one is using a cell-free system, and this may not be free from artifacts. We must therefore turn to techniques using intact cells. The main information we have here is from mutagenesis, and this has already

proved very valuable by confirming many features of the code, as well as suggesting that the code is probably universal. It must be used with discretion since, for various technical reasons, it is occasionally possible to get a mistaken result. However, mutagenesis never tells us the identity of a particular triplet but merely the relationships *between* triplets, and the same is true of the results obtained by the phase-shift method. The latter is especially useful since it is often possible to decipher unambiguously what the base sequence must have been. Thus it is possible not only to establish the meaning of a certain triplet, but to show that it was in fact used in an actual message. This can only be done by mutagenesis in special cases, such as those involving the *ochre* and *amber* triplets, and also possibly methionine and tryptophan if it turns out, as seems likely, that each of these has only a single codon. Unfortunately, it appears very difficult to produce enough examples of phase-shifts to make it a method which can be used very widely.

Finally, we have the interesting possibility of confirming the code by studying the anticodons on the various transfer RNA molecules—a possibility first considered about 1958. At the moment this looks very promising. Holley originally suggested (Holley et al., 1965) that the triplet IGC at positions 36 to 38 in the nucleotide sequence of yeast alanine tRNA might be the anticodon, and this has been strongly supported by the sequences recently worked out for two serine tRNAs and one tyrosine tRNA, both from yeast. Unfortunately, the work needed to find the nucleotide sequence of a large number of tRNA molecules is very considerable, especially since some of them may be present in rather small amounts. Moreover, we must be sure of the rules for pairing between codon and anticodon before we can interpret the results. I have suggested (Crick, 1966) what these rules should be, and they are set out in Table 1. The present experi-

mental evidence, though supporting them, is a long way from proving them.

Although we think we know where the anticodon lies on the tRNA, it is still very uncertain which part of the tRNA is recognized by its activating enzyme. In particular, it is controversial whether the enzyme interacts at all with the anticodon itself. Personally, I would suspect the region rich in dihydrouridylic acid to be the main one involved.

It is also not going to be easy without single-crystal X-ray work on crystals of tRNA (if they can be obtained) to be sure of the way tRNA is folded. However, it is a reasonable hypothesis that the secondary and tertiary structures of different tRNA molecules have many common features, and it may be possible to deduce the folding of different tRNA molecules by an imaginative comparison of their base sequences. The sequences already available appear to suggest that the "four-leafed clover" model, suggested (among others) by Holley, may be a good approximation to the secondary structure (Holley et al., 1965).

Whatever the structure of tRNA may be, it is unlikely to be a very simple one. It almost appears as if tRNA were Nature's attempt to make an RNA molecule play the role of a protein. Looked at in this way, the many unusual bases that tRNA contains make good sense. It may be very difficult to stabilize an intricately folded molecule using only the four standard bases, which is all that can be incorporated by the normal replication process.

Then there are the punctuation marks to consider. We are certain that the *amber* triplet (UAG) terminates the polypeptide chain in *E. coli* infected with phage T4, and all the evidence suggests that it does so in uninfected *E. coli*. Recent genetic work makes it likely that the *ochre* triplet (UAA), when it occurs as a mutant, also terminates the chain. We certainly suspect that the *amber* and *ochre* triplets, especially the latter, are used for natural chain termination, but this has yet to be proved. Nor do we yet know the chemical mechanism for releasing the polypeptide chain.

The situation for chain initiation is also somewhat obscure. We know that in *E. coli* formyl methionine is often involved in chain initiation, and we know something about the relevant triplets, and can make a good guess at the mechanism. Whether there are other mechanisms for chain initiation in *E. coli* is uncertain. There is much that needs to be tidied up here, and we are at the moment unclear how chain initiation occurs in higher organisms. On the present evidence it seems that in *E. coli* a triplet such as GUG stands for one amino acid, namely methionine, when it initiates a chain, and another amino acid, namely valine, when it is in the middle of a chain.

TABLE 1.

Anticodon	Codon
U	A ⎱ G ⎰
C	G
A	U
G	U ⎱ C ⎰
I	U ⎱ C ⎬ A ⎰

Predicted pairing rules for the third position of the codon, according to Crick (1966).

One of the results of this work on chain initiation has been to direct attention to the correct concentration of divalent cations, especially Mg^{++}, needed for *accurate* protein synthesis in the cell-free system. If a particular mRNA does not include a chain-initiating triplet, it can only start by making a mistake. This is more likely to happen at higher Mg^{++} concentrations, and thus the "optimum" value for Mg^{++} may be artificially high, and liable to cause other coding errors.

Nonsense mutants bring us to the question of polarity—the effect of a mutation in one gene on the expression of a neighboring gene. It is quite certain that some nonsense mutants are polar, but the degree of polarity seems to vary with the position of the mutation and with the operon being considered. Our recent knowledge of chain-initiating triplets may help us to explain some of these anomalies.

The skeleton in our cupboard is the possibility of ambiguity. It may be that even in the middle of a message a certain triplet may stand for more than one amino acid. Of course, we know that this can happen when mistakes occur in the translation mechanism, either due to so-called suppressor genes or due to the influence of various small molecules, such as certain antibiotics. What is at issue, however, is whether in a *normal* cell there is any ambiguity of translation. If there is, it will make establishment of the code much more difficult.

Even when we know the genetic code, we will still not know what it is that signals the beginning of a gene and the end of a gene, nor the beginning of an operon and the end of an operon. In particular, we do not know whether these particular base-sequences have any special relationship to the genetic code proper. It seems to me that the whole question of regulation is at the present in a confused state. It is certainly not clear, for example, whether tRNA is normally involved in regulating the *rate* of protein synthesis in any important way.

However, all these topics, which are covered at some length in the various papers in this symposium, have at least one thing in common—it should be possible within the next few years to obtain precise and unambiguous answers to them, using the experimental techniques that we have available. The same, unfortunately, cannot be said about the other remaining major problem, which was dealt with in the last session of this meeting. That concerns the structure of the genetic code and its origin. The difficulty here, as I see it, is that it is not going to be easy to produce any evidence that will decide definitively between the various ideas that people are beginning to put forward. Whereas in the Fifties we had to endure a whole lot of rather

poor papers on the nature of the genetic code, in the last years there has been a rash of papers on its structure and origin. I am considering offering an annual prize for the worst paper published on this subject—I don't think there will ever be any lack of candidates for it.

Not all the ideas so far suggested have been bad. It is obvious, for example, that the hydrophobic amino acids tend to cluster on the left-hand side of the table of the genetic code, and that the charged amino acids group together to some extent on the right-hand side. People have been showing great eagerness in trying to explain this, although nobody has investigated whether the present grouping is likely to have happened by chance: that is, whether by putting reasonable restrictions on the allocations of triplets and drawing them out of a hat, one would often find *some* grouping or other which looked significant.

The main point at issue, however, is whether the code has a structural basis, or whether it mainly arose by chance. If the former were true there should be some stereo-chemical relationship between each amino acid and the triplets that code it, or with the corresponding anticodons. So far this has not been demonstrated. On the other hand, the adaptor hypothesis, in its extreme form, implies that no such stereo-chemical relationships need exist. We are thus compelled to consider the origin and structure of tRNA, and the early mechanism for protein synthesis.

The matter is discussed more fully in the paper by Woese and his collaborators. The point I want to emphasize here is that we may be heading for a very unhealthy situation, in that theory will run far ahead of useful experimental facts. One of the reasons that I enumerated, in this introduction, something of the early history of the code was to show how little theory was able to contribute. In particular, it seems highly likely that the comma-less code would be widely accepted today if we did not have such good experimental evidence against it, simply because it can be derived in a very elegant manner from a rather sensible postulate. I hope, therefore, that when people put forward detailed theories about the origin of the genetic code, they will try if possible to produce ones which can be tested in some way or other.

Meanwhile we are bound to pass through a period when general ideas on the subject have to be aired. For example, it is an obvious speculation that although the mechanism always moved along three bases at a time in reading the genetic message, the original code may have recognized the first two bases and the rather wobbly recognition of the last one may have been a later development. Equally, anyone can see that the original nucleic acid may

only have had two bases instead of four, and it is quite amusing to discuss which two it might have been. A possibility suggested by Leslie Orgel, which is not quite so obvious, is that in the beginning the two bases were adenine and inosine. But none of these ideas will get us very far unless we can find some way of obtaining more experimental evidence, either frozen in the present organisms or from dramatic experimental results. If somebody could show in several cases that there really was a stereo-chemical affinity between a certain triplet and a certain amino acid, and that this correlated with the genetic code, I would be more impressed by such experimental evidence than by any amount of theoretical argument or model building.

But putting all these doubts and reservations to one side, one can say, looking at the papers in this symposium, that the elucidation of the genetic code is indeed a great achievement. It is, in a sense, the key to molecular biology because it shows how the two great polymer languages, the nucleic acid language and the protein language, are linked together. It is not only important to know the details for their own sake, but by knowing these details we become quite confident that our general ideas, such as the sequence hypothesis, are indeed correct. It will be difficult, after this, for doubters not to accept the fundamental assumptions of molecular biology which we have been trying to prove for so many years.

REFERENCES

BEADLE, G. M. 1945. Biochemical genetics. Chem. Rev. *37:* 15–96.

BELOZERSKY, A. N., and A. S. SPIRIN. 1958. A correlation between the compositions of deoxyribonucleic and ribonucleic acids. Nature *182:* 111–112.

BRACHET, J. 1944. Embryologie chimique. Masson et Cie., Paris. 510p.

BRENNER, S. 1957. On the impossibility of all overlapping triplet codes in information transfer from nucleic acid to proteins. Proc. Natl. Acad. Sci. *43:* 687–694.

CALDWELL, P. C., and C. HINSHELWOOD. 1950. Some considerations on autosynthesis in bacteria. J. Chem. Soc. *4:* 3156–3159.

CASPERSSON, T. 1947. The relations between nucleic acid and protein synthesis. Symp. Soc. Exp. Biol. *1:* 127–151.

CRICK, F. H. C. 1957. Discussion, p. 25–26. *In:* The structure of nucleic acids and their role in protein synthesis. Biochemical Soc. Symp. no. 14, Cambridge Univ. Press.

——. 1958. On protein synthesis. Symp. Soc. Exp. Biol. *12:* 138–163.

——. 1966. Codon-anticodon pairing: The wobble hypothesis. J. Mol. Biol., *19:* 548–555.

CRICK, F. H. C., L. BARNETT, S. BRENNER, and R. J. WATTS-TOBIN. 1961. General nature of the genetic code for proteins. Nature *192:* 1227–1232.

CRICK, F. H. C., J. S. GRIFFITH, and L. E. ORGEL. 1957. Codes without commas. Proc. Natl. Acad. Sci. *43:* 416–421.

DOUNCE, A. L. 1952. Duplicating mechanism for peptide chain and nucleic acid synthesis. Enzymologia *15:* 251–258.

ECK, R. 1963. Genetic code: Emergence of a symmetrical pattern. Science *140:* 477–481.

GAMOW, G. 1954. Possible relation between deoxyribonucleic acid and protein structure. Nature *173:* 318.

——. 1954. Possible mathematical relation between deoxyribonucleic acid and protein. Kgl. Danske Videnskab. Selskab. Biol. Medd. *22:* no. 3. p. 1–13.

HOAGLAND, M. 1960. The relationship of nucleic acid and protein synthesis as revealed by studies in cell-free systems, p. 349–408. *In* E. Chargaff and J. N. Davidson [ed.] The nucleic acids, vol. III. Academic Press, New York.

HOLLEY, R. W., J. APGAR, G. A. EVERETT, J. T. MADISON, M. MARQUISSE, S. H. MERRILL, J. R. PENSWICK, and A. ZAMIR. 1965. Structure of a ribonucleic acid. Science *147:* 1462–1465.

LEDER, P., and M. NIRENBERG. 1964. RNA codewords and protein synthesis. II. Nucleotide sequence of a valine RNA codeword. Proc. Natl. Acad. Sci. *52:* 420–427.

LEE, K. Y., R. WAHL, and E. BARBU. 1956. Contenu en bases puriques et pyrymidiques des acides désoxyribonucléiques des bactéries. Ann. Inst. Pasteur *91:* 212–224.

NIRENBERG, M., and J. MATTHAEI. 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. Proc. Natl. Acad. Sci. *47:* 1588–1602.

RYCHLIK, I., and F. ŠORM. 1962. Replacements of amino acids in proteins and ribonucleic acid coding. Collection Czechoslov. Chem. Commun. *27:* 2686–2691.

SARABHAI, A. S., A. O. W. STRETTON, S. BRENNER, and A. BOLLE. 1964. Co-linearity of the gene with the polypeptide chain. Nature *201:* 13–17.

SINSHEIMER, R. 1959. Is the nucleic acid message in a two-symbol code? J. Mol. Biol. *1:* 218–220.

YANOFSKY, C., B. C. CARLTON, J. R. GUEST, D. R. HELINSKI, and U. HENNING. 1964. On the colinearity of gene structure and protein structure. Proc. Natl. Acad. Sci. *51:* 266–272.